| Title: | Machine Learning using local environment descriptors to predict new scintillator materials |
|---|---|
| Author(s): | Williams, Logan Douglas<br>Pilania, Ghanshyam |
| Intended for: | Group meeting talk at University of Michigan – Ann Arbor |
| Issued: | 2018-09-10 |

# Machine Learning using local environment descriptors to predict new scintillator materials

Logan Williams[1], Ghanshyam Pilania[2]

[1]University of Michigan – Ann Arbor
[2]Los Alamos National Laboratory

# Scintillation Counters

- Scintillator + Photodetector
- Scintillator converts ionizing radiation to visible light
- Photodetector converts light to electrical signal
  - Photomultiplier tube
  - Photodiode
  - CCD (Charge-Coupled Device)
- Detect and measure ionizing radiation
  - Different materials good for detecting:
    - Alpha
    - Beta
    - Gamma
    - X-rays
    - Neutrons
  - Can measure radiation intensity and energy

# Scintillator Counter Applications

- Medical imaging
- Scientific research
  - High energy physics experiments
  - Astrophysics
  - Nuclear stockpile research
    - E.g. DARHT uses scintillators with CCDs for high resolution and speed
    - Also: pRAD and the upcoming ECSE and MaRIE facilities
- Nuclear worker safety
- Detection of nuclear material
  - Nonproliferation
  - Homeland security
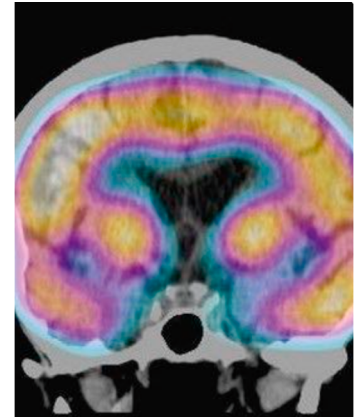
# Scintillation Process

- **Energy Conversion**
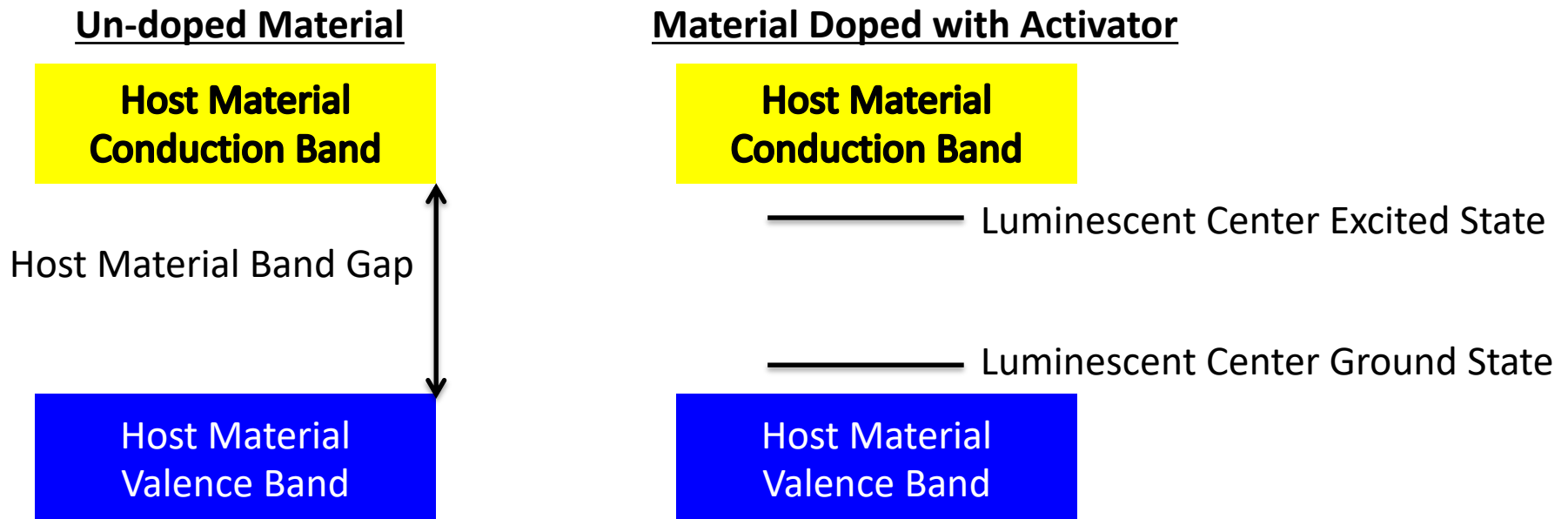  - Ionizing radiation converted to form many "hot" electrons and holes

- **Thermalization**
  - Electrons and holes undergo inelastic scattering
  - Lose energy to the crystal as heat and relax to the band edges

- **Localization**
  - Electrons and holes localize as excited states at luminescent centers

- **Light Emission**
  - Electron and hole recombine at a luminescent center, emitting light

# Material Influence on Scintillation

- **Energy Conversion**
  - Host crystal structure determines efficiency of energy conversion
    - High electron density materials efficiently capture gamma radiation
    - Materials rich in hydrogen efficiently capture neutrons

- **Thermalization**
  - The valence and conduction band dispersions control the mobility of the thermalizing holes and electrons, respectively

- **Localization**
  - Capture rate is determined by the energy difference between the VBM and luminescent center's ground state for holes, and the difference between the CBM and the luminescent center's excited state for electrons

- **Light Emission**
  - Rate of recombination determined by the luminescent center used

# Luminescent Center Requirements

- States must be within the band gap to capture electrons and holes
- States should be close to the valence and conduction bands for fast capture
- States should not be extremely close or thermal excitation could de-localize captured charge carriers

**Un-doped Material**

**Material Doped with Activator**

| Host Material Conduction Band |
|---|

Host Material Band Gap

| Host Material Valence Band |
|---|

| Host Material Conduction Band |
|---|

——————— Luminescent Center Excited State

——————— Luminescent Center Ground State

| Host Material Valence Band |
|---|

# Ce$^{3+}$ as a Luminescent Center

- Lanthanides, such as Ce$^{3+}$, are commonly used activators
- Energy levels relative to host can vary tremendously



Possibly useable with band gap engineering
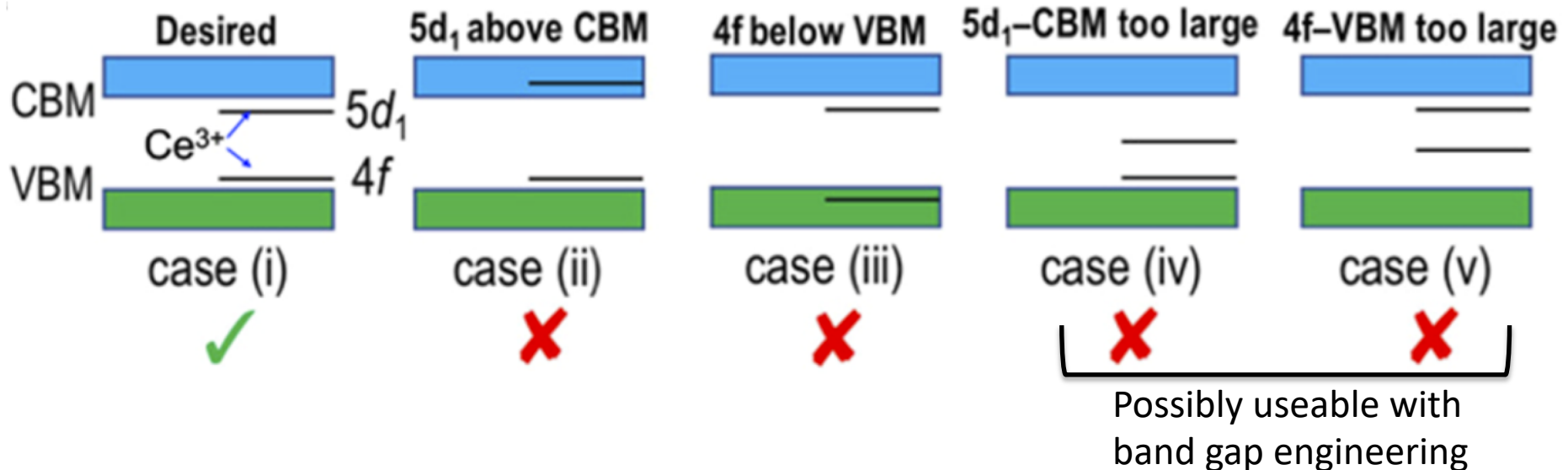
# Goal: Scintillator Discovery

- Develop a way to screen an incredibly wide chemical space
- Discover new scintillators quickly and easily



ML Predictions of $Ce^{+3}$ $4f$ and $5d_1$ levels for $A_2(B,B')X_6$ Elpasolites

Down-select

Image from G. Pilania, K.J. McClellan, C.R. Stanek, and B.P. Uberuaga, J. Chem. Phys. **148**, 241729 (2018).
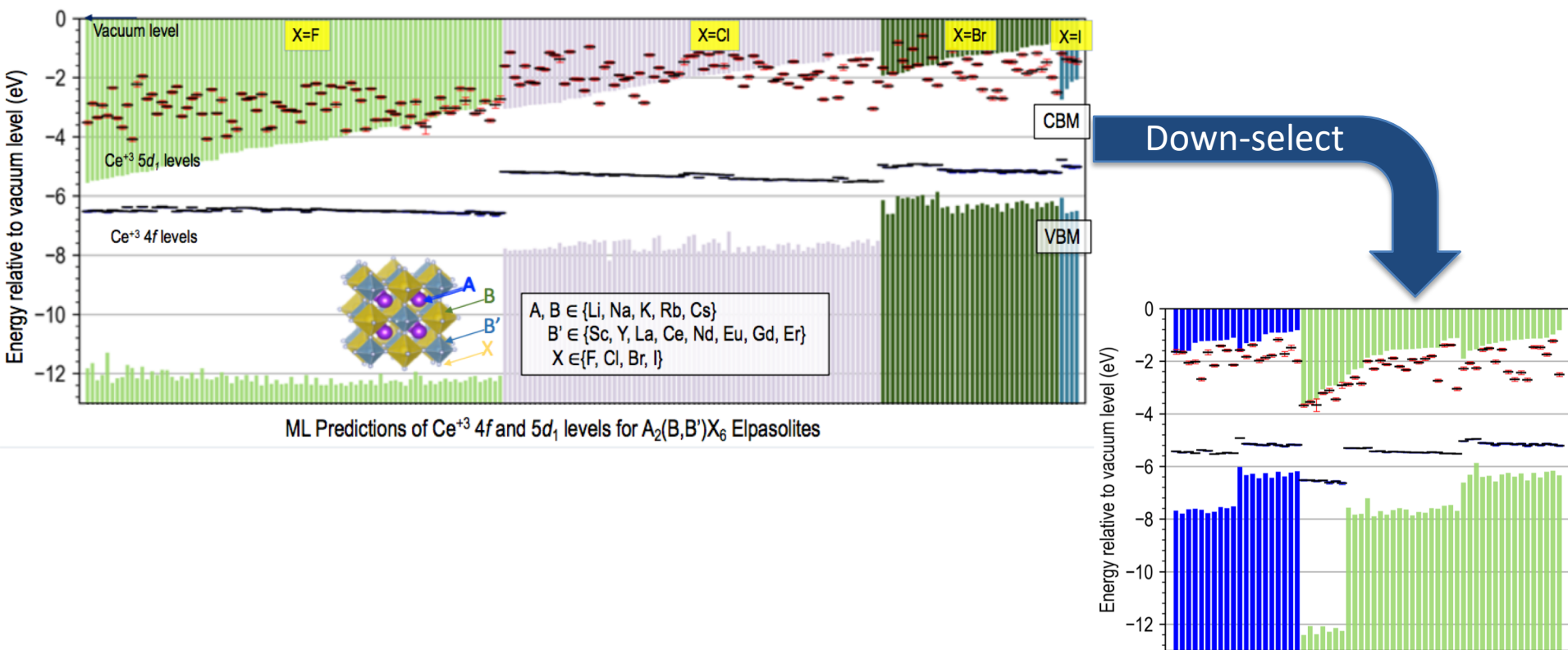
# Data-Driven Science Paradigm

- Discovery of novel scintillators has mostly been through trial-and-error experimentation guided by chemical intuition.

- Computational scintillator prediction is difficult.

- Machine learning can be used to discover new scintillator materials.

Right image from C. Draxl and M. Scheffler, **2**, 1 (2018), arXiv: 1805.05039
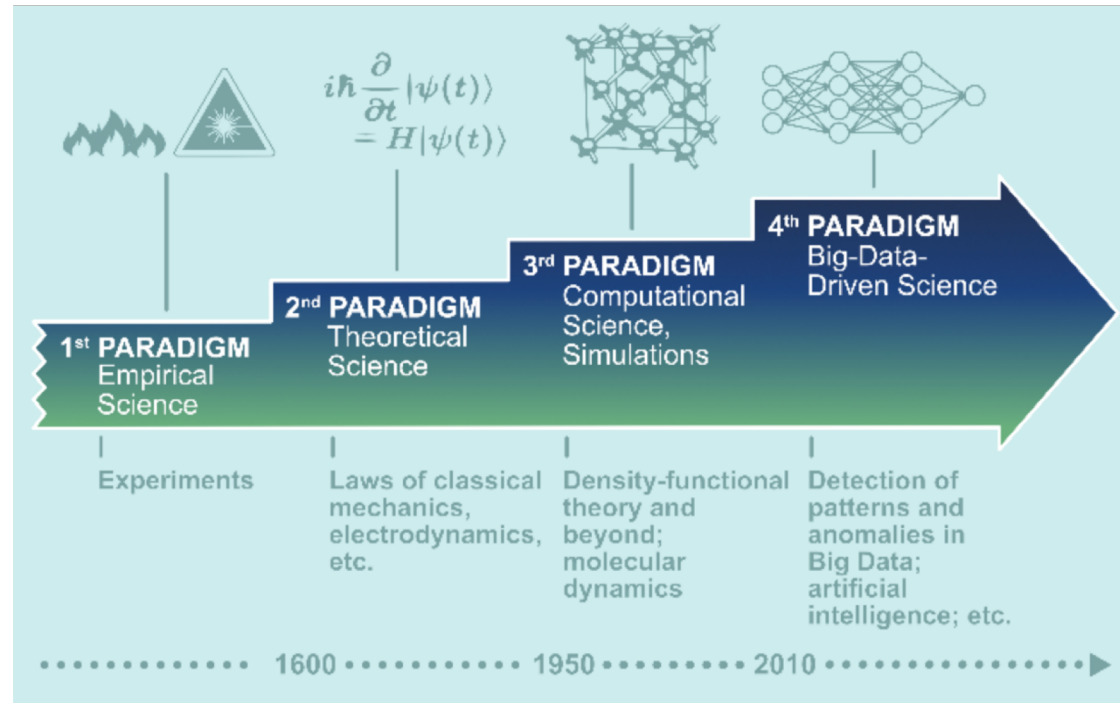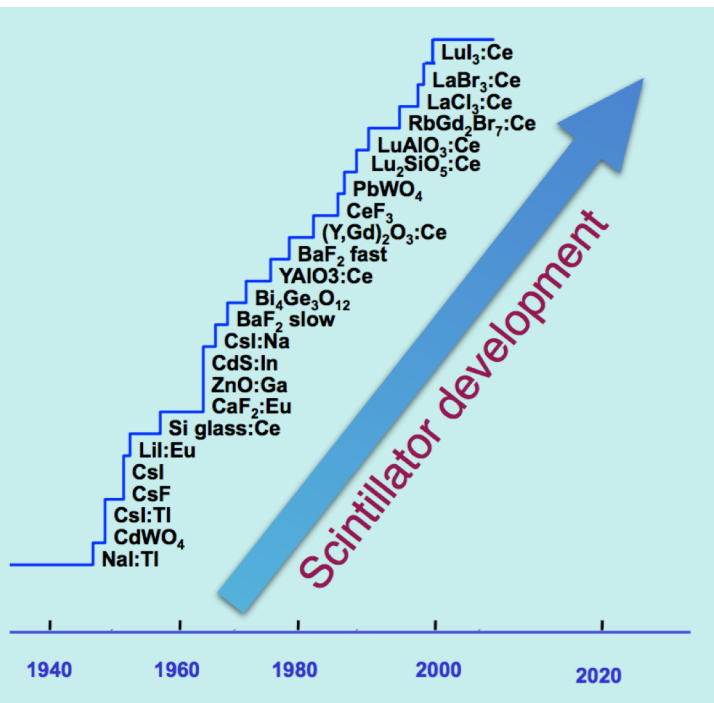
# Predicting Ce$^{3+}$ energy levels

- Two useful parameters for predicting Ce$^{3+}$ 4f and 5d$_1$ energy levels:
- For 4f energy levels: U(6, A) – the 4f electron binding energy difference between Eu$^{2+}$ and Eu$^{3+}$



Image from G. Pilania, K.J. McClellan, C.R. Stanek, and B.P. Uberuaga, J. Chem. Phys. **148**, 241729 (2018).
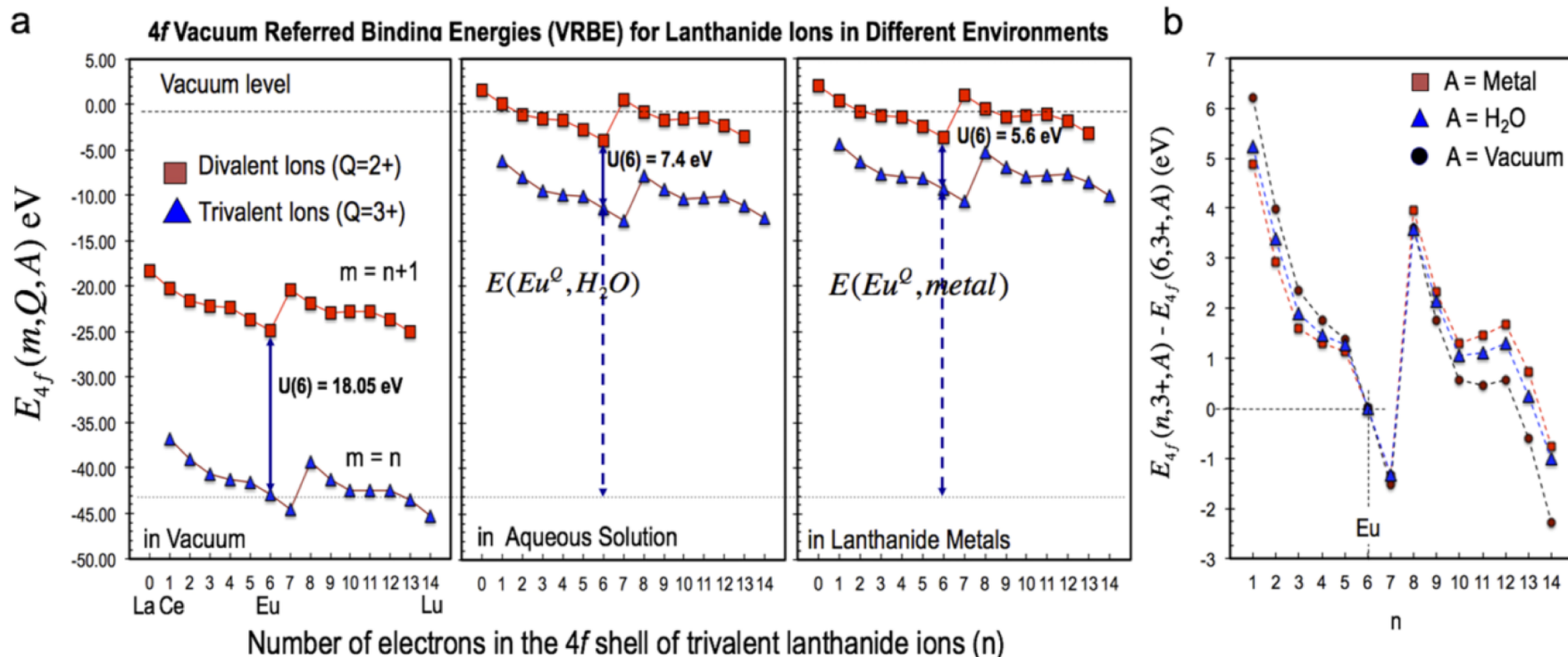
# Predicting Ce³⁺ 4f energy levels

- The chemical shift, $E(Eu^Q, A)$, directly depends on $U(6,A)$.

$$E_{4f}(7, 2+, A) = -24.92 + \frac{18.05 - U(6,A)}{0.777 - 0.0353\, U(6,A)},$$

$$E_{4f}(6, 3+, A) = E_{4f}(7, 2+, A) - U(6,A).$$

- $E(Eu^Q, A) = E_{4f}(6, 3+, A) - E_{4f}(6, 3+, vacuum)$



$$E_{4f}(m, Q, A) = E_{4f}(m, Q, vacuum) + E(Eu^Q, A)$$

$$+ f\,\frac{E(Eu^Q, A)^2}{1440Q}\Delta R(m).$$

Images from G. Pilania, K.J. McClellan, C.R. Stanek, and B.P. Uberuaga, J. Chem. Phys. **148**, 241729 (2018).
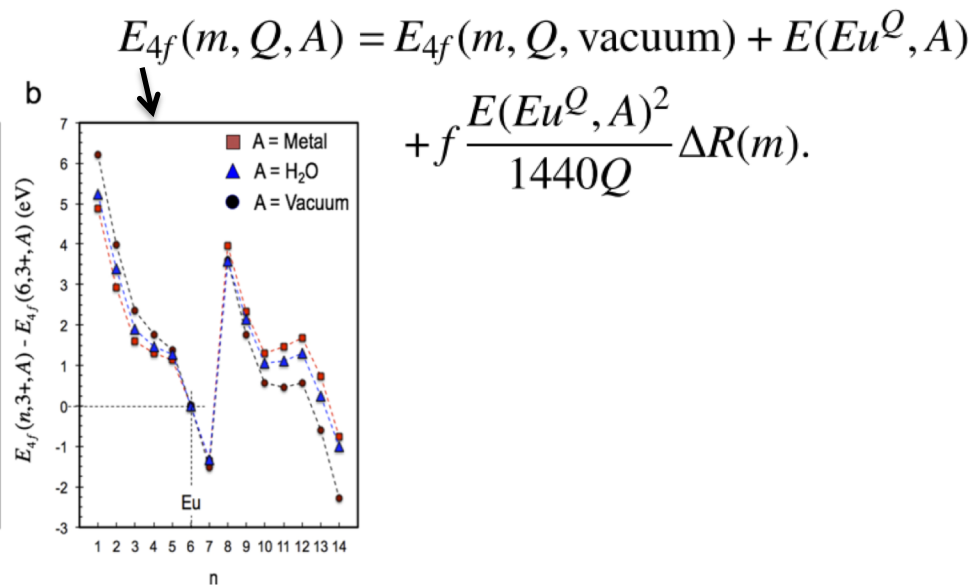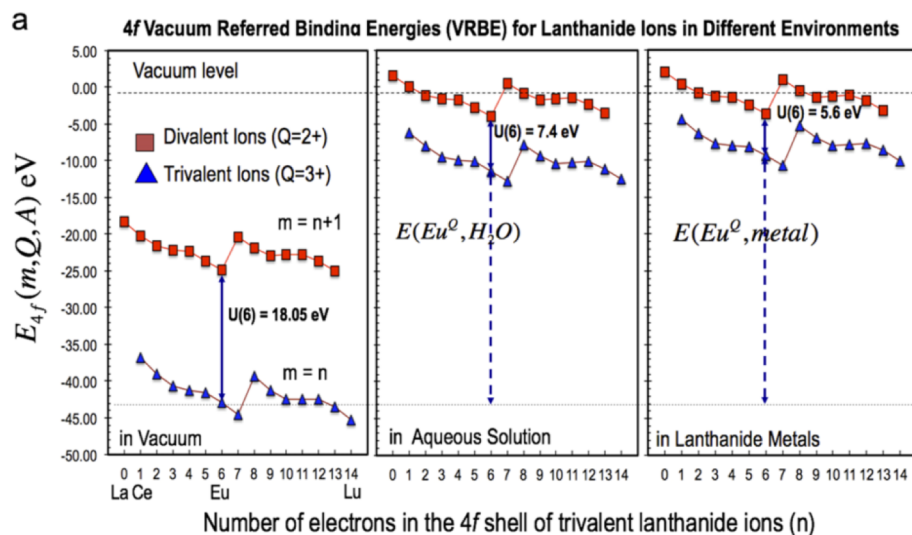
# Predicting Ce$^{3+}$ 5d energy levels

- For 5d energy levels: D(Ce$^{3+}$, A) – the crystal field depression or spectroscopic red shift

- The change in the energy difference between the 4f and lowest 5d energy levels of Ce$^{3+}$ when going from vacuum to a host environment

- Highly dependent upon local environment!
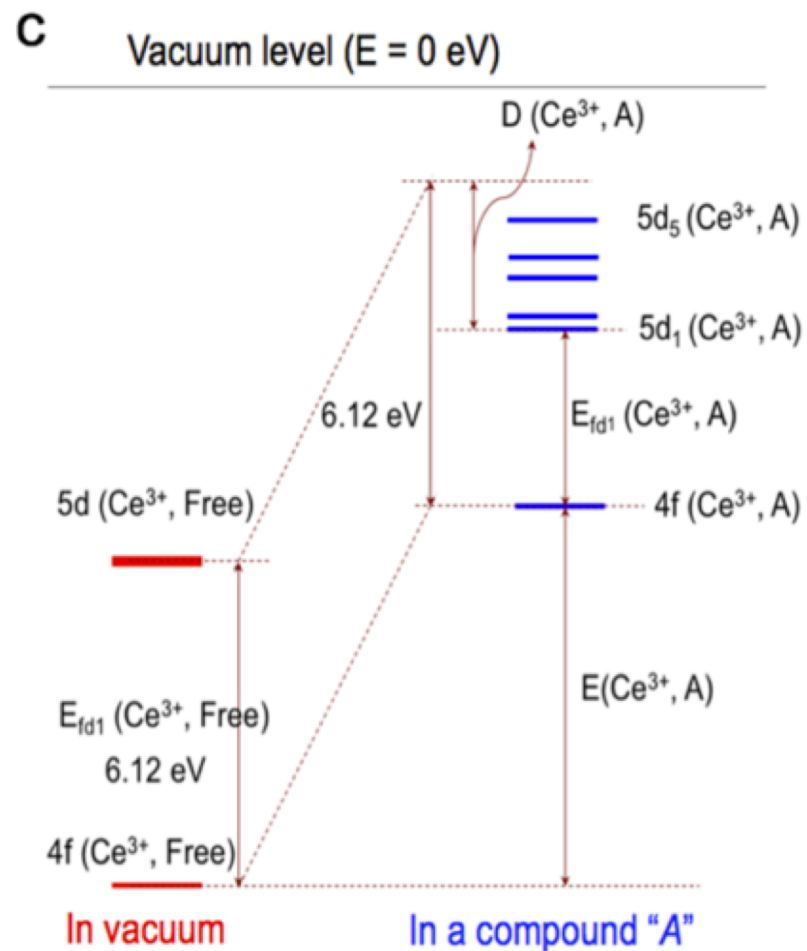
# Materials Datasets

- Using two separate datasets that contain the U and D values for a set of compounds that exist in the Materials Project database.

P. Dorenbos, Journal of Luminescence 135, 93 (2013)
P. Dorenbos, Journal of Luminescence 91, 155 (2000)

A. Jain et al. Apl Materials 1, 011002 (2013)
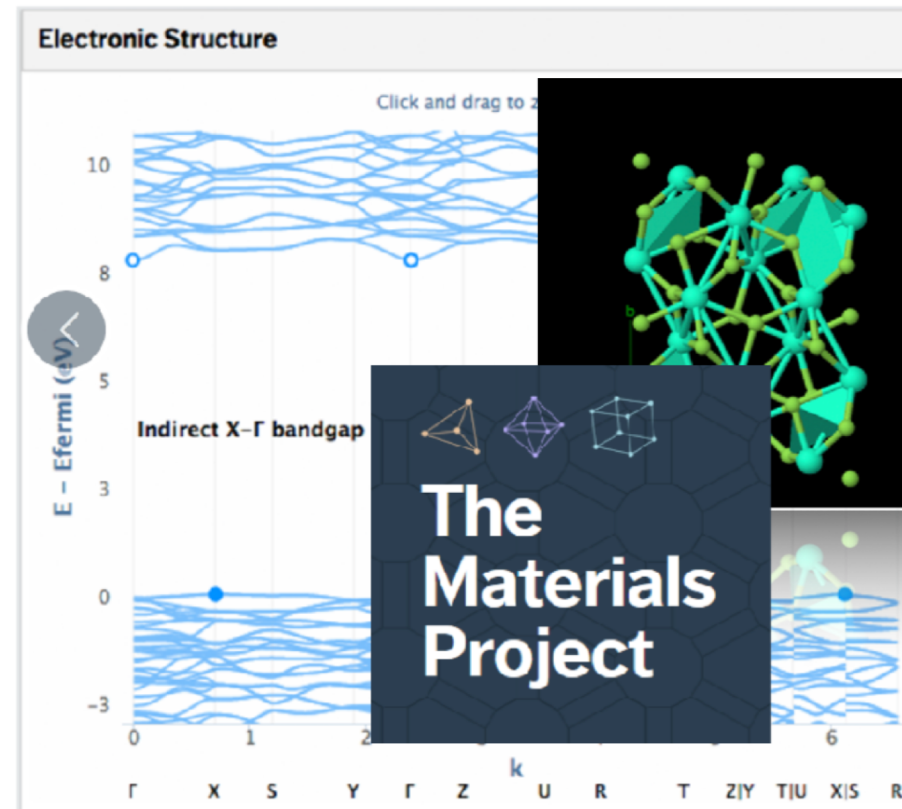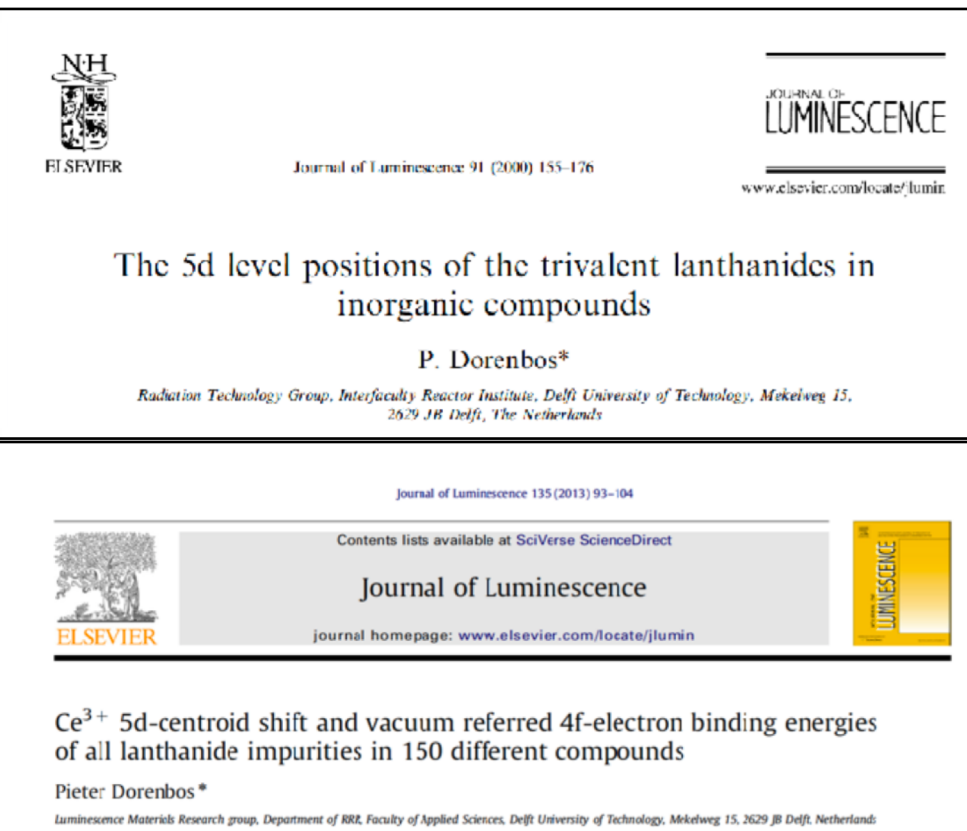
# Machine Learning Algorithm

- We use Kernel Ridge Regression (KRR)
- Kernel Ridge Regression is a modification of least-squares fitting
  - The kernel is a function that can be applied to each pair of inputs to calculate inner products of those inputs in a higher-dimensional feature space
    - Allows for tractable computation of nonlinear relationships in high or infinite dimensional space
    - We use a Gaussian kernel in our calculations
  - The 'ridge' is the addition of a penalty function that depends on the magnitude of the fitting coefficients
    - Counteracts the tendency to 'overfit' the data to perfectly match the training data

Model predictions:

$$f(\tilde{\boldsymbol{x}}) = \sum_{i=1}^{n} \alpha_i k(\boldsymbol{x}_i, \tilde{\boldsymbol{x}})$$

Optimization:

$$\arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^n} \sum_{i=1}^{n} \left(f(\boldsymbol{x}_i) - y_i\right)^2 + \lambda \left\|\boldsymbol{\alpha}\right\|_{\mathcal{H}}^2$$

$$\Leftrightarrow \quad \arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^n} \langle \boldsymbol{K}\boldsymbol{x} - \boldsymbol{y}, \boldsymbol{K}\boldsymbol{x} - \boldsymbol{y} \rangle + \lambda \boldsymbol{\alpha}^T \boldsymbol{K} \boldsymbol{\alpha}$$

Solving by setting gradient to zero:

$$\nabla_{\boldsymbol{\alpha}} \ \boldsymbol{\alpha}^T \boldsymbol{K}^2 \boldsymbol{\alpha} - 2\boldsymbol{\alpha}^T \boldsymbol{K}\boldsymbol{y} + \boldsymbol{y}^T \boldsymbol{y} + \lambda \boldsymbol{\alpha}^T \boldsymbol{K} \boldsymbol{\alpha} = 0 \quad \Leftrightarrow \quad \boldsymbol{K}^2 \boldsymbol{\alpha} + \lambda \boldsymbol{K} \boldsymbol{\alpha} = \boldsymbol{K}\boldsymbol{y}$$

$$\Leftrightarrow \quad \boldsymbol{\alpha} = (\boldsymbol{K} + \lambda \boldsymbol{I})^{-1} \boldsymbol{y}.$$

13

# Current Status - descriptors

| Feature ID | Symbol | Description | Feature ID | Symbol | Description |
|---|---|---|---|---|---|
| 1 | $E_g$ | Bandgap (computed within DFT using the PBE functional ) | 11 | $EA_{av}$ | Average electron affinity |
| | | | 12 | $EA(+)_{av}$ | Average electron affinity for cationic species |
| 2 | $\rho$ | Density of host compound | | | |
| 3 | $\chi_{av}$ | Average electronegativity | 13 | $EA(-)_{av}$ | Average electron affinity for anionic species |
| 4 | $\chi(+)_{av}$ | Average electronegativity for cationic species | 14 | $r(+)_{av}$ | Average empirical radius for cationic species |
| 5 | $\chi(-)_{av}$ | Average electronegativity for anionic species | 15 | $r(-)_{av}$ | Average empirical radius for anionic species |
| 6 | $\chi(+)_{av}^{q}$ | Charge weighted average electronegativity for cationic species | 16 | $P_{av}$ | Average atomic polarization |
| 7 | $\chi(-)_{av}^{q}$ | Charge weighted average electronegativity for anionic species | 17 | $P(+)_{av}$ | Average atomic polarization for cationic species |
| 8 | $IP_{av}$ | Average ionization potential | 18 | $P(-)_{av}$ | Average atomic polarization for anionic species |
| 9 | $IP(+)_{av}$ | Average ionization potential for cationic species | 19 | $M(+)_{av}$ | Average Pettifor's Mendeleev's number for cationic species |
| 10 | $IP(-)_{av}$ | Average ionization potential for anionic species | 20 | $M(-)_{av}$ | Average Pettifor's Mendeleev's number for anionic species |

# Current Status - results

- U(6,A) is well described – 4f level are easy to predict
- D(Ce$^{3+}$, A) is difficult – 5d levels are sensitive to local environment



Image from G. Pilania, K.J. McClellan, C.R. Stanek, and B.P. Uberuaga, J. Chem. Phys. **148**, 241729 (2018).
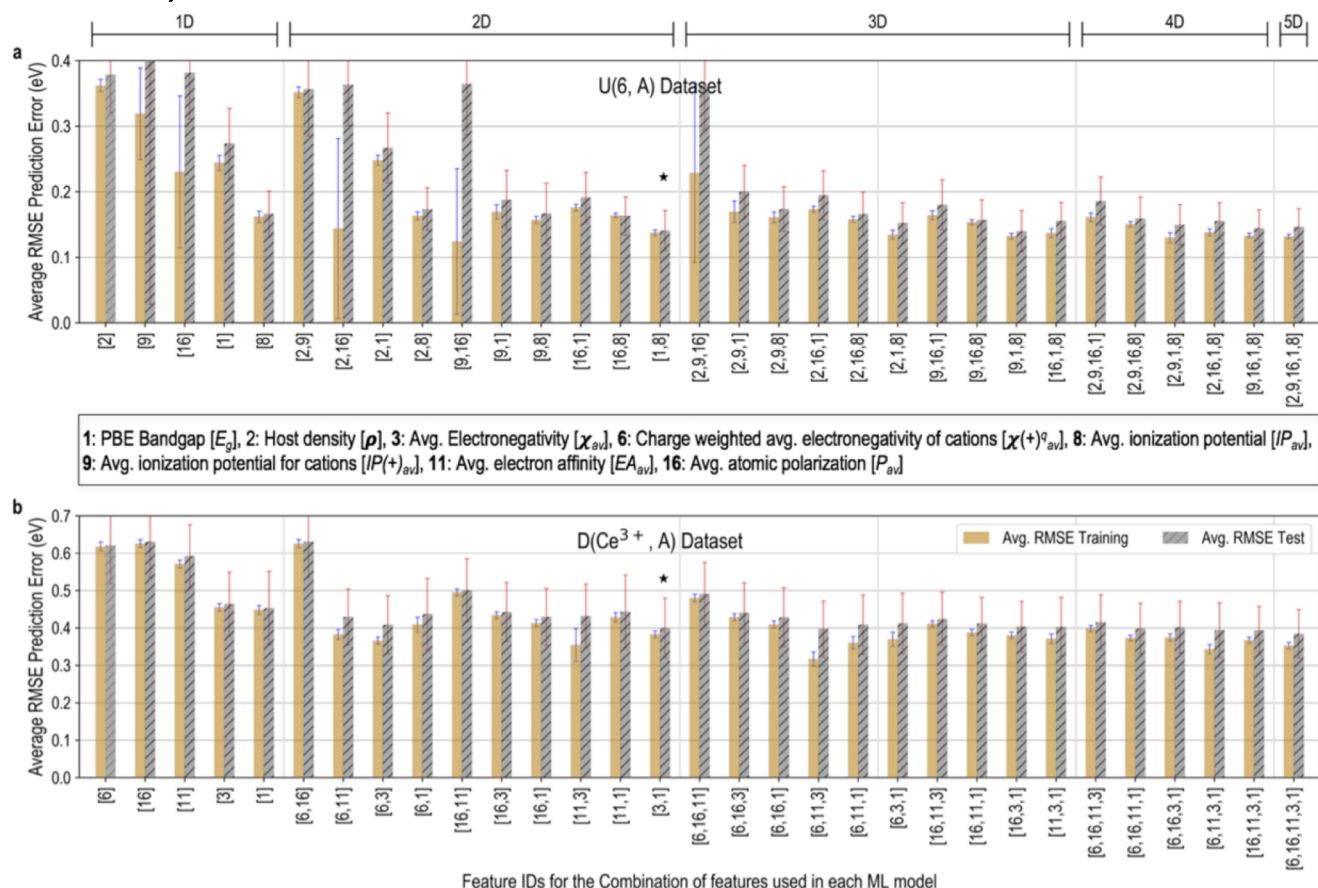
# Doping Site Selection

- To describe the local environment, we first need to predict what site the $Ce^{3+}$ dopant will occupy.

- Assumptions made:
  - Any atomic site is a possible doping site.
  - $Ce^{3+}$ will not occupy any interstitial sites or form complex defect clusters.

- Methodology:
  - Assign a penalty to ionic radius mismatch between $Ce^{3+}$ and the site's normal ion.
  - Assume that a site being too small for $Ce^{3+}$ is 3x worse than being too large.
  - Assign a penalty to charge state difference between $Ce^{3+}$ and the site's normal ion.
  - Assume a difference in charge state of 1 is equivalent to an ionic radius difference of 0.2 Å (where the site is too large).
  - Select the site in the structure with the smallest penalty as the doping site.

# Local Environment Descriptors

- **Gaussian Radial Distribution Function (RDF)**
  - Radial description of surrounding neighbors
- **AGNI Fingerprints**
  - Product integral of RDF and Gaussian window functions, angular dependence
- **Gaussian Symmetry functions**
  - Based on pair distances and angles, separated into radial and angular functions
- Voronoi Fingerprint
  - Voronoi-tessellation based features
    - Volume, area, and nearest-neighbor statistics
    - Indices (number of i-edged facets), i-fold symmetry indices
- Gaussian Angular Fourier Series (AFS)
  - Product of distance functionals and cosine of the angle of a pair from the origin
- Coordination Number
- Order Parameter Site Fingerprint
  - Locates neighbor shells for a range of coordination numbers
- Local Difference in Electronegativity
- **Crystal Field Splitting Spread** – from a Point Charge model

$5d_5$ ($Ce^{3+}$, A)

$5d_1$ ($Ce^{3+}$, A)
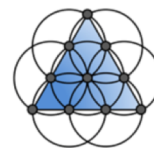
# Local Descriptor Results

- AGNI Fingerprints and Gaussian Symmetry functions both encode local environment details – notably angular relationships



**Feature IDs:**

0: original two
1: AGNI fingerprints (default)
2: AGNI fingerprints (optimized)
3: Gaussian Symmetry functions (default)
4: Gaussian Symmetry functions (optimized)
5: crystal field splitting spread (current best)

6: Gaussian RDF
7: Voronoi Fingerprint
8: Gaussian AFS
9: Coordination Number
10: OP Site Fingerprint
11: local difference in electronegativity
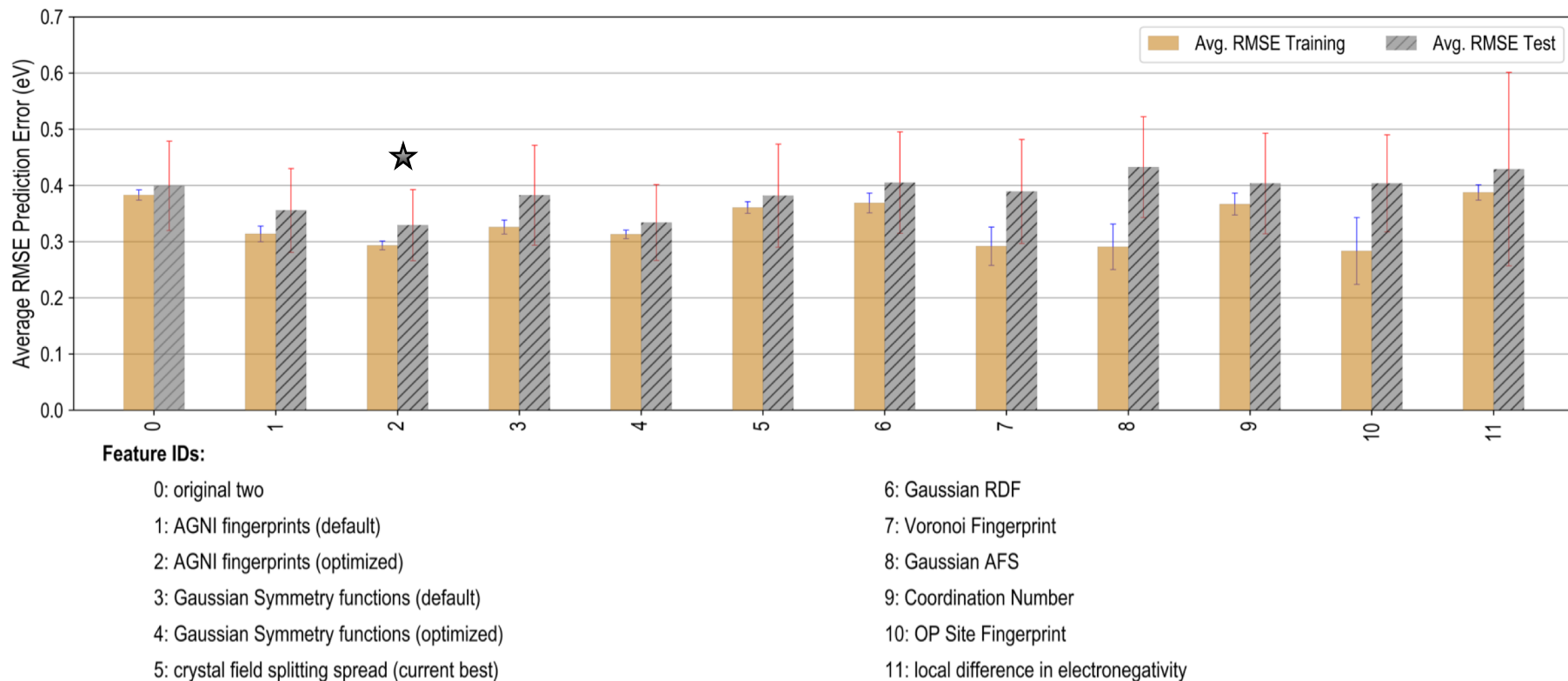
# AGNI and Gaussian optimization

- Reducing the featurized distance to 3.75 Å produces optimal results
- Crystal field splitting is controlled by nearest neighbor interactions



AGNI

Gaussian Symmetry

Avg. RMSE Training    Avg. RMSE Test

Average RMSE Prediction Error (eV)

Feature IDs:

0: AGNI fingerprints d=10.0 Angstroms

1: AGNI fingerprints d=8.0 Angstroms

2: AGNI fingerprints d=6.0 Angstroms

3: AGNI fingerprints d=5.0 Angstroms

4: AGNI fingerprints d=4.0 Angstroms

5: AGNI fingerprints d=3.75 Angstroms

6: dividor

7: Gaussian Symmetry functions d=6.5 Angstroms

8: Gaussian Symmetry functions d=4.0 Angstroms

9: Gaussian Symmetry functions d=3.75 Angstroms

10: Gaussian Symmetry functions d=3.5 Angstroms

# Issues with the Crystal Field Splitting Descriptor

- Crystal Field Spectrum Spread ≠ Crystal Field Depression

- Crystal Field Spectrum Spread is highly sensitive to local atomic arrangement
  - Doping with $Ce^{3+}$ can cause very large local rearrangement around the $Ce^{3+}$ doping site

  - Possible to detect cases where this will be a problem using doping site selection criterion?



C

Vacuum level (E = 0 eV)

D ($Ce^{3+}$, A)

$5d_5$ ($Ce^{3+}$, A)

Crystal Field Spectrum Spread

$5d_1$ ($Ce^{3+}$, A)

6.12 eV

$E_{fd1}$ ($Ce^{3+}$, A)

5d ($Ce^{3+}$, Free)

4f ($Ce^{3+}$, A)

$E_{fd1}$ ($Ce^{3+}$, Free)

6.12 eV

E($Ce^{3+}$, A)

4f ($Ce^{3+}$, Free)
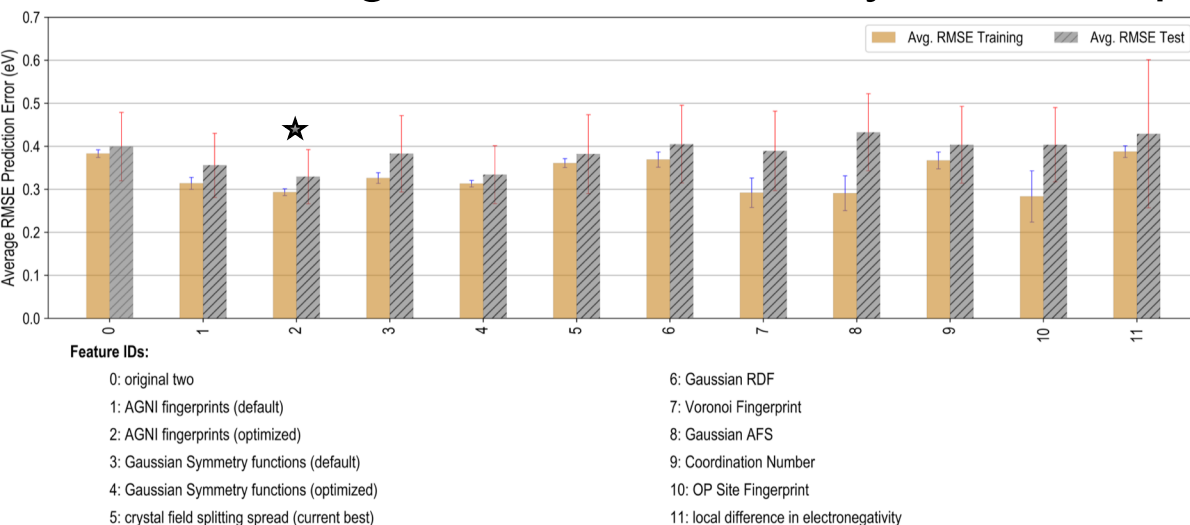
In vacuum

In a compound "A"
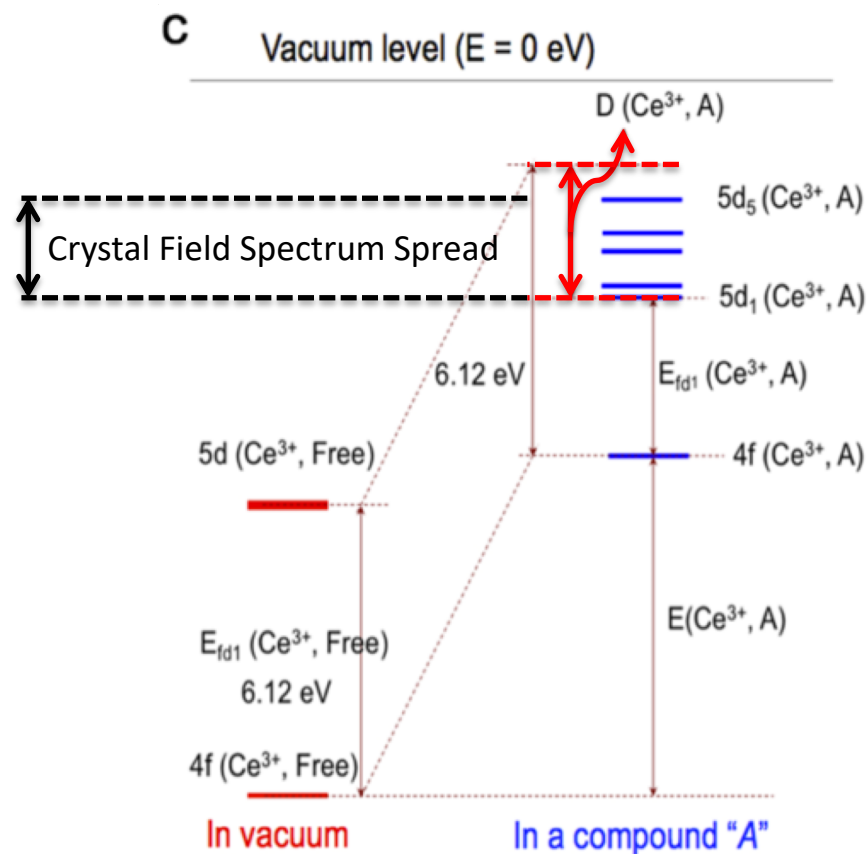
# Summary, Acknowledgements, and Future Work

Local Environment descriptors can improve prediction of $D(Ce^{3+},A)$, and thus 5d energy levels. [0.39 -> 0.33 eV RMSE]

Future improvements could come from:

- Improved doping site selection function
- Optimized descriptor settings
- Pruning of database for crystal field splitting



Feature IDs:
0: original two
1: AGNI fingerprints (default)
2: AGNI fingerprints (optimized)
3: Gaussian Symmetry functions (default)
4: Gaussian Symmetry functions (optimized)
5: crystal field splitting spread (current best)
6: Gaussian RDF
7: Voronoi Fingerprint
8: Gaussian AFS
9: Coordination Number
10: OP Site Fingerprint
11: local difference in electronegativity
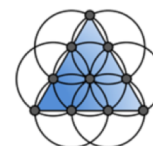
# Extra Slides

# Kernel Ridge Regression Overview

- **Regression is a form of supervised learning**
  - Start with a training set of input and corresponding outputs
  - Predict the output of inputs not in the training set

- **Learning a finite system has infinite solutions**
  - Assumptions made to choose the 'best' model
  - Simpler is better
    - Smoothness, or regularization, used to find simpler models

# Linear Regression

- Simplest form of regression
- Fitting function is in the shape: $f(\tilde{\boldsymbol{x}}) = \sum_{i=1}^{d} \beta_i \tilde{x}_i = \langle \boldsymbol{\beta}, \tilde{\boldsymbol{x}} \rangle$

- Fitting is done by minimizing the squared error in the training set:
$$\underset{\boldsymbol{\beta} \in \mathbb{R}^d}{\arg\min} \sum_{i=1}^{n} \left( \langle \boldsymbol{\beta}, \boldsymbol{x}_i \rangle - y_i \right)^2$$

- Optimization done by setting the gradient to zero:
$$\nabla_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left( \langle \boldsymbol{\beta}, \boldsymbol{x}_i \rangle - y_i \right)^2 = \boldsymbol{0} \quad \Leftrightarrow \quad \boldsymbol{\beta} = \left( \boldsymbol{X}^T \boldsymbol{X} \right)^{-1} \boldsymbol{X}^T \boldsymbol{y}$$

# Ridge regression

- Linear regression is prone to overfitting errors
  - Large coefficients $\beta_i$ that cancel in the training set
  - In the prediction of new inputs, these cause large errors
- Ridge regression adds regularization to prevent overfitting
  - Penalty term added to the optimization problem:

$$\underset{\boldsymbol{\beta}\in\mathbb{R}^d}{\arg\min} \sum_{i=1}^{n}\left(\langle\boldsymbol{\beta}, \boldsymbol{x}_i\rangle - y_i\right)^2 + \lambda\left\|\boldsymbol{\beta}\right\|^2$$

- Solving the optimization problem gives:

$$\boldsymbol{\beta} = \left(\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{I}\right)^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

# Kernel ridge regression

- Kernel ridge regression adds non-linearity
- Maps inputs into a higher dimensional space and applies the linear algorithm there
  - Mapping into higher dimensional space correspondingly increases computational complexity – how to avoid?
- The kernel trick
  - ML algorithms can be written to only use inner products between inputs
  - Kernel functions operate on input space vectors, but yield the same results as inner product evaluations in feature space
- The Gaussian kernel is widely applicable:

$$k(\boldsymbol{x}, \boldsymbol{z}) = \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{z}\|^2}{2\sigma^2}\right)$$

26

Equations taken from M. Rupp, Int. J. Quantum Chem. **115**, 1058 (2015).

# Kernel ridge regression

- Fitting function is in the shape:   $f(\tilde{\boldsymbol{x}}) = \sum_{i=1}^{n} \alpha_i k(\boldsymbol{x}_i, \tilde{\boldsymbol{x}})$

- Fitting is done by minimizing the squared error in the training set plus the regularization penalty:

$$\underset{\boldsymbol{\alpha} \in \mathbb{R}^n}{\arg\min} \sum_{i=1}^{n} \left( f(\boldsymbol{x}_i) - y_i \right)^2 + \lambda \left\| f \right\|_{\mathcal{H}}^2$$

$$\Leftrightarrow \quad \underset{\boldsymbol{\alpha} \in \mathbb{R}^n}{\arg\min} \left\langle \boldsymbol{K}\boldsymbol{x} - \boldsymbol{y}, \boldsymbol{K}\boldsymbol{x} - \boldsymbol{y} \right\rangle + \lambda \boldsymbol{\alpha}^T \boldsymbol{K} \boldsymbol{\alpha}$$

- Optimization done by setting the gradient to zero:

$$\nabla_{\boldsymbol{\alpha}} \ \boldsymbol{\alpha}^T \boldsymbol{K}^2 \boldsymbol{\alpha} - 2\boldsymbol{\alpha}^T \boldsymbol{K}\boldsymbol{y} + \boldsymbol{y}^T \boldsymbol{y} + \lambda \boldsymbol{\alpha}^T \boldsymbol{K} \boldsymbol{\alpha} = \boldsymbol{0} \quad \Leftrightarrow \quad \boldsymbol{K}^2 \boldsymbol{\alpha} + \lambda \boldsymbol{K} \boldsymbol{\alpha} = \boldsymbol{K}\boldsymbol{y}$$

$$\Leftrightarrow \quad \boldsymbol{\alpha} = (\boldsymbol{K} + \lambda \boldsymbol{I})^{-1} \boldsymbol{y}.$$